Table of Contents

November - December 1994 -- Volume 2, Number 6

IBM SP2 Delivers Performance, Challenges to Consortium Partners

by Elisabeth Wechsler

Initial configurations of the IBM SP2 system passed acceptance tests and the NASA Computational Aerosciences (CAS)-funded Cooperative Research Agreement (CRA) remains on schedule. However, significant challenges await the consortium in the months ahead for both hardware and software development, according to various members of the project team at NAS and IBM.

NAS personnel have expressed considerable satisfaction with the performance that has been achieved to date (*see sidebar*). Nonetheless, some problems remain. The principal area of concern is system software, including the need for dynamic job management and parallel file systems support. IBM and NAS Systems Division personnel are looking into these areas on behalf of the consortium.

For background information about the CRA, see the *July-August issue of* NAS News.

"We expected some shortcomings in system software and hope to work with IBM to make the system more usable to researchers," said David Bailey, who coordinates the CRA research project at NAS.

Richie Jacobovits, IBM program manager for the CRA, said: "We've seen the benefits of collaboration and IBM is really happy with the way things are going." He continued, "The implementation of the acceptance test was much more rigorous than we expected and we're glad we've been able to live up to the expectations NASA has set." He explained that NAS "created a script that ran the evaluation test suite around the clock. The acceptance test was very much a stress test because the test script ran the SP2 at peak performance rates for nearly two weeks."

Making a `True' Collaboration

"We're trying to make this a true collaboration," said Toby Harness, of the NAS parallel systems group. "Our goal is to make IBM do well -- there's no `vendor' involved because it's a CRA."

Chuck Niggley, of the NAS parallel systems science support group, echoed that point of view: "We've established a rapport within various IBM development groups so we can have nondisclosure access to some of their future design specifications. They are beginning to welcome our input."

Each week, 15-18 NAS staff and researchers attend a teleconference, joined by 2-3 IBM employees working on site at NAS, and hooked up with IBM engineers and managers from research labs throughout the country.

Installation and Tests Offered Surprises

The original SP2 system, consisting of 64 nodes and named "babbage," was delivered to NAS on July 5. Acceptance testing (coordinated by Bernard Traversat, NAS parallel systems group) began a few days after installation and was completed on July 31. On July 25, the second 64-node system arrived. Called "cabbage," it passed the post-shipment test on August 11 and its acceptance test on September 8. This phase was coordinated by David McNab, also of the NAS parallel systems support group.

Babbage passed its acceptance test on the first eligible day, based on the CRA timetable. Cabbage "took a little longer because it was shipped with new versions of the CPU card," which experienced some early manufacturing problems, according to Jacobovits. These problems have since been resolved and babbage's CPU cards were likewise expected to be upgraded during October.

With the delivery of an extra switch frame at the end of September, the two 64 node systems have now been connected as a single 128-node system, still called "babbage." The 128-node system performs on a par with a 16-processor CRAY C90," Bailey said.

An additional 32 nodes arrived in October and will be connected with the 128 nodes to create a 160-node system. It is hoped that the 160-node system will be fully available to CAS researchers and other users by January, Bailey continued.

In addition to assisting with the acceptance tests, members of the NAS systems support group helped install approximately 300 user accounts on the system. They are also devising workarounds for problems as they arise, so that users can run their code on the SP2. There are plans to install NAS-developed programs such as NTV, AIMS, and PBS, as well as IBM software such as LoadLeveler and PVMe (a tuned implementation of PVM) on the SP2.

Portability Issues Pose Challenge

"We've also been involved in dealing with IBM to get the research version of their message-passing interface (MPI) implementation, so that users will have it available," Niggley said. He added that "portability issues are driving our efforts to get MPI on the SP2 as soon as possible."

"We're doing a lot of pioneer work to figure out how to minimize the amount of disk space needed by the operating system," Niggley continued. His group plans to test a beta version of IBM's high-performance Fortran (HPF) implementation, which is similar to other versions of HPF and is expected to save researchers the chore of "writing explicit message-passing calls in their code."

Niggley's group has begun to use HyperText Markup Language (HTML) to "get our information out there as fast as possible to help new users." He added that there is some urgency, given that the Intel iPSC/860 supercomputer was removed from NAS at the end of September.

One difficulty of disseminating information to users is that IBM's online documentation system, InfoExplorer, cannot be customized for the NAS system because the files are written in a proprietary format, Niggley said.

System Software is Major Focus

Harness and his group are focused on making the system software work on the SP2. In particular, they are heading development efforts for a dynamic job scheduler, parallel file systems (PFS) project, and accounting software for tracking projects and monitoring resource consumption.

A dynamic job and process scheduler is one that must respond to changes in system workload and resources, intervening where necessary, Harness explained. To achieve this goal, the group must find replacements for the Network Queuing System (NQS) after evaluating other queuing systems. In the long run, there will be one product on the SP2; in the interim, the group plans to install both IBM's LoadLeveler product, the new NAS Portable Batch System (PBS), and possibly others.

"Dynamic scheduling and automatic load balancing are crucial for massively parallel processor (MPP) systems," Jacobovits said. These are being investigated at IBM's T. J. Watson Research Center in Yorktown Heights, NY, as part of the collaboration, he said.

The parallel file system (PFS) project is expected to provide parallel file abstractions so a system like the SP2 can perform I/O in a "natural" manner--that is, several files acting like one so that the researcher doesn't need to make adjustments manually, Harness said.

Extending Program Functionality

To accomplish this goal, Harness's group is getting early versions of PFS and Vesta (its precursor version) from IBM sites. The plan is to extend the functionality of PFS, specify and write the programmer interface for MPI I/O, and then work with IBM to do an implementation of MPI I/O for the SP2.

The SP2 is helping NAS meet its goal of providing MPI on all its systems, by serving as a testbed for MPI development, according to Harness. "MPI looks like it will be the standard for message passing, but currently I/O needs to be written for each machine, so MPI I/O will be a standard interface for I/O," he said, adding that he hopes the development will be completed by March.

The accounting software is expected to be installed on the SP2 in December, Harness said. "The

accounting system is an NPSN [NAS Processing System Network]-wide project, which was started before the SP2 was installed but in anticipation of it. It's been written from the ground up."

He added that "in a workstation-style system like the SP2, a lot of the administrative tools don't scale well beyond a few tens or hundreds of nodes. And many tools are still missing from the system." Harness commented, "These problems were somewhat anticipated, but they've turned out to be more significant than first realized. It's potentially a very big challenge."

IBM Sees Future in MPP Systems

When asked how IBM viewed the CRA contributing to its long-term business strategy, Jacobovits focused on MPP systems technology. He believes the way for IBM to achieve this goal is to "use the microprocessor and operating system from our RISC System/6000, add a highly scalable packet switch (designed to scale to 32,000 nodes), and develop parallel system software to requirements for robustness and reliability typically found in our commercial mainframe computers."

SP2 Demo At Supercomputing '94

IBM plans to demonstrate a 64-node SP2 in its booth at Supercomputing '94, November 14-18, using the so-called airplane code authored by Tony Jameson, principal investigator at Intelligent Aerodynamics Inc., Princeton, NJ-- another CRA partner. Jacobovits is also trying to organize a birds-of-a-feather session for all CRA research partners, since many will attend the conference.

For information about doing SP2 research outside the scope of the High Performance Computing and Communications Program, send email to NAS User Interface Manager Pat Elson at **pelson@nas.nasa.gov**.

The newest parallel supercomputer system at NAS, a 128-node IBM SP2 configuration called "babbage," consists of 16 towers of 8 nodes each. The system was installed in September and serves approximately 300 user accounts. The three-year Cooperative Research Agreement is a partnership with industry, government, and university representatives.

Next Article | Contents | Main Menu | NAS Home

Early SP2 Results `Encouraging'

by Elisabeth Wechsler

Early performance results for the IBM SP2 have been encouraging.

On September 15, researchers at Centric Engineering Systems Inc., a Cooperative Research Agreement (CRA) partner in Santa Clara, CA, reported achieving performance of 2.5 gigaflops on 32 nodes of the SP2, according to David Bailey of NAS. (*See the table below on NAS Parallel Benchmark (NPB) results for the IBM SP2.*)

Benchmark	SP2/128	C90/16
EP	9.60	11.80
MG	2.63	3.97
CG	26.79	10.61
FT	14.49	8.43
IS	1.99	0.98
LU	49.83	51.60
SP	63.80	80.40
BT	75.41	96.40

A comparison of the IBM SP2 and the 16-processor CRAY C90, based on results of the 128-node test. These results, obtained in September, show that the SP2 is faster on five of the eight NAS Parallel Benchmarks, including all three CFD applications (LU, SP, and BT).

According to Leonardo Dagum, who supervises NPB testing and analysis, all three CFD benchmarks -- from a total of eight -- "are indicative of the types of actual data movement and computation required in state-of-the-art CFD application codes."

Bailey added that "we believe the NPB are fairly important and reliable benchmarks." Because of the publicity afforded the NPB results in the supercomputing community, he believes the comparisons between vendors are realistic because the performance "is the best they can make it." He emphasized that initial performance on a user application might be only 10 million floating-point operations per second (MFlop/s) per node. But with a reasonable tuning effort, users are seeing results of 40 to 80 MFlop/s per node.

The latest NPB results are regularly posted to the World Wide Web.

Next Article | Contents | Main Menu | NAS Home

MPIRUN: A Loader for Multidisciplinary and Multizonal MPI Applications

by Samuel Fineberg

MPIRUN, a NAS-developed utility, allows users to create multidisciplinary and multizonal application programs that are portable to any parallel computer or to clusters of workstations. MPIRUN works with both C and Fortran and is currently available on all NAS parallel systems.

Programs Important in Aerosciences

Multidisciplinary and multizonal applications are an important class of programs in the area of computational aerosciences. In these codes, two or more distinct parallel applications or copies of a single application are used to model a single problem. These parallel applications are usually implemented using message passing. All message-passing libraries support simple applications consisting of a single program running on many processors (often called SPMD). However, support is lacking for applications where different programs are running on each processor. Unfortunately, this is the most straightforward way to implement multidisciplinary and multizonal applications, utilizing a separate program for each discipline or zone.

In the past, these applications were implemented using extensions to vendor supplied message-passing libraries; for example, the intercube library for the Intel iPSC/860 and the Map library for the Intel Paragon. However, programs written with these libraries are not portable.

There are many portable message-passing libraries, some of which support the communications requirements of multidisciplinary and multizonal applications. One well-known portable message-passing library is Parallel Virtual Machine (PVM) from the University of Tennessee; another is the new Message Passing Interface standard (MPI) developed by the Message Passing Interface Forum, a group of researchers and developers from universities, government research laboratories (including NAS), and industry.

While PVM is sufficient for creating multidisciplinary and multizonal applications, it lacks support for fast communication operations involving sets of processors (that is, collective communication) and efficient creation of new processor sets. This means that programs suffer a performance penalty for using PVM instead of the vendor-supplied message-passing library. Also, PVM is not a standard, and therefore, it changes from time to time.

MPI does support fast collective communication and processor set creation. In addition, MPI does not impose any inherent performance penalties on programs. For example, IBM's T. J. Watson Research Center developed a version of MPI for the SP2 (called MPI-F), which is actually faster than IBM's proprietary message passing layer (MPL).

Can Be Used to Develop MPI Programs

The problem with MPI is that it only specifies message passing; there is currently no standard for loading MPI applications. The NAS parallel systems group developed MPIRUN to resolve this problem. MPIRUN is a portable program loader that can be used to develop MPI programs. By using MPIRUN, it is possible to create complete multidisciplinary and multizonal applications that can be run on any parallel system.

MPIRUN is available on the Intel Paragon, Thinking Machines CM-5, and IBM SP2. It also runs on the Intel iPSC/860, workstation clusters, and the IBM SP1. For more detailed information on MPI and MPIRUN, use lynx or Mosaic to view the following URL addresses:

http://parallel.nas.nasa.gov/Parallel/Tools/how_to_use_mpi.html
http://parallel.nas.nasa.gov/Parallel/Tools/how_to_use_mpirun.html

To request a copy of the MPI standard, send email to **doc-center@nas.nasa.gov**. Or, enter **help mpirun** on NAS parallel systems lovelace, babbage, or grace.

Next Article | Contents | Main Menu | NAS Home

NASA Ames Seeks Proposals for Scientific Visualization

The 1995 NASA Research Announcement (NRA), a solicitation for research proposals in scientific visualization, was issued August 15 and announced in *Commerce Business Daily* by NASA Ames Research Center. The NRA is funded by the NAS Systems Division and awards grants to scientists for research that will support NAS objectives. The intent of the grants is to foster technology transfer for the scientific visualization of computational fluid dynamics and related aerosciences, particularly multidisciplinary simulations of unsteady flow over complex geometries.

Deadline: November 14

Although proposals may be submitted at any time, those received by November 14 will be considered for a 1995 award. Proposals received after the closing date may still be considered if the selecting official considers the proposal to offer NASA a significant technical advantage or cost reduction. Single-year awards of up to $100,000 per year, which also allow researchers access to required computational resources, are considered. Total funding for this year will be approximately $400,000. Multiple-year proposals (up to three years and renewed annually) are also accepted.

Specific areas in which the NRA seeks to fund outside research encompass visualization methods, user interfaces, quantitative validation and benchmarking, and other related scientific visualization technologies. The NRA document provides more specifics.

Current research funded under NRA grants includes such topics as visualization techniques and rapid prototyping environments. Summaries of grants awarded under the 1993 and 1994 NRAs are available on the World Wide Web at the following URL: http://www.nas.nasa.gov/NAS/Grants/.

Announcement Copies Required

To submit a research proposal, scientists must first request a required copy of the announcement from: Lupe Velasquez, Contracting Officer, NASA Ames Research Center, Mail Stop 258-5, Moffett Field, CA 94035-1000, or send a fax request to (415) 604-4377.

Address technical questions about the areas of research supported under the NRA to Jeff Hultquist at (415) 604-4970, or send email to **hultquist@nas.nasa.gov**.

For information on the NRA program, send email to Arsi Vaziri at **vaziri@nas.nasa.gov**.

Next Article | Contents | Main Menu | NAS Home

Langley Researchers Develop Novel Technique to Investigate Laminar Turbulent Transition

by C. David Pruett and Chau-Lyan Chang

Predicting and controlling laminar-turbulent transition--a fundamental change experienced by the flow of air close to the surface of an aircraft--is of major importance to designers in the aerospace industry. A novel technique that combines two existing computational approaches for investigating transition has been recently developed by a team of research scientists contracted to NASA Langley Research Center, Hampton, VA.

Laminar-turbulent transition is a complex physical process by which the flow of a fluid (including air) over a surface changes from an orderly layered "laminar" state to a chaotic "turbulent" state. (The sudden change in the stream of smoke rising from a cigarette in a room of still air is one common example of laminar-turbulent transition.)

Transition prediction is especially important in the design of supersonic and hypersonic flight vehicles, such as the High-Speed Civil Transport (HSCT) and the National AeroSpace Plane. However, because the physics of transitioning high-speed flows is not well understood--in part because stability and transition experiments are difficult to perform when the Mach number, a measure of speed relative to the speed of sound, is high--computation has assumed a particularly important role.

Increasingly, researchers are using computers for "direct numerical simulation" (DNS) of transition, a process by which the Navier-Stokes equations, which govern the motions of fluids, are solved numerically. Even on supercomputers, DNS is a daunting task because many millions of grid points are required to ensure that both large-scale and small-scale motions of the fluid are accurately represented.

A Grand Challenge Problem

Despite decades of cooperative research among scientists in the fields of stability theory, physical experimentation, and numerical computation, predicting and controlling laminar-turbulent transition has remained elusive. Even before prediction and control of transition will become practical, scientists must gain further insight into the process, a task for which DNS is well suited. Thus, simulating the transition of a high-speed flow has become one of the computational grand challenges of the '90s.

Several difficulties have previously thwarted the attainment of this goal; a particular stumbling block is that, because transition occurs gradually at high speeds, the computational domain for a simulation is necessarily long. Fortunately, recent advances in theory and computational science have made this problem approachable. Among these advances are: the development of the parabolized stability equation (PSE) method by Bertolotti, *et al.*; the adaptation of high-order numerical methods to DNS; and the availability of a class of supercomputers large enough to handle the problem, such as the NAS CRAY C90, with a gigabyte of main memory.

Accuracy Important to Vehicle Design

Scientists believe that laminar-turbulent transition originates when small oscillations in the fluid become "unstable"; that is, initially small fluctuations begin to grow, eventually attaining large (nonlinear) amplitudes that in turn trigger a final "breakdown" to turbulent flow.

Because both the drag (resistance) and the frictional heating experienced by aerospace vehicles increase dramatically as the air transitions from a laminar to a turbulent state, it is important for vehicle designers to be able to accurately predict transition onset and the extent of the transition region. Otherwise, the vehicle may be "overdesigned" and fail to meet performance or efficiency standards--or "underdesigned," in which case safety is a concern. In addition, there are potentially dramatic economic gains for aircraft manufacturers who can realize the improved vehicle efficiency associated with a delay of transition onset by an emerging technology known as laminar-flow control.

Approach Combines Methods

During the past year, researchers at NASA Langley simulated the onset of transition in Mach 8 flow along a sharp cone of a circular cross section (a configuration of interest because it resembles, for example, the shapes of vehicle forebodies or engine inlet cones). Their simulation was unique in several ways:

1. For reasons of computational efficiency, most previous DNS models assumed that the flow evolved in time rather than in space, although it is generally accepted that the spatial-evolution model is more physically correct. The Langley computation was the first spatial DNS of a high-speed flow to attain transition onset (as indicated by a rapid increase in drag).

2. The simulation was also among the first to extend DNS methodology to a configuration of practical interest (that is, the cone). Most previous simulations have considered only idealized configurations (such as, flow along a flat plate).

3. Langley's approach involved a synergistic combination of PSE and DNS methodologies.

Briefly, the PSE method approximates the Navier-Stokes equations, making possible a very efficient

solution procedure. Langley researchers used the PSE method to identify likely transition mechanisms and to compute the earlier stages of the transition process, in which the fluctuations in the flow have low to moderately large amplitudes. The breakdown stage of transition, in which fluctuations are extremely large, was then computed by DNS using inflow boundary conditions derived from the PSE method.

In DNS, the full Navier-Stokes equations are solved without approximation-- although at great computational expense. The advantage of combining PSE and DNS was that the size of the domain could be limited for the expensive (DNS) part of the computation.

The DNS algorithm incorporates several recent advances in computational technology. Specifically, a high-order, low-storage Runge-Kutta method was used for time advancement, whereas both spectral and high-order compact-difference schemes were used for spatial discretizations.

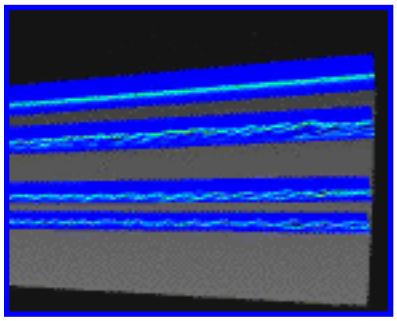
In addition to being highly accurate, the DNS algorithm was optimized to facilitate autotasking on multiple processors of vonneumann, the NAS CRAY C90; a 2.5 gigaflop rate was attained on eight CPUs. The final run for the DNS required some 27 million grid points and used nearly 512 megawords of main memory. In total, the simulation required in excess of 1000 C90 CPU hours.

Collaboration May Solve `Mysteries'

The mysteries of laminar-turbulent transition may ultimately be solved if scientists in the fields of experimentation, theory, and computational fluid dynamics join forces on well-defined "building-block" problems. In that spirit, the Langley computation was patterned as much as possible on a wind-tunnel experiment of Stetson and co-workers. Al-though the computational and experimental results differed in many important aspects, and there are now many new questions to be answered, the Langley DNS results are invaluable in several ways:

- PSE results were shown to agree with DNS until near the point of transition onset, raising hopes that the PSE method can be used as a transition-prediction tool. Since most PSE calculations can be done in a matter of minutes, in contrast to hundreds of hours for DNS, the importance of validating the PSE method for applications to high-speed flow cannot be overstated.

- Previously unexplained phenomena observed in wind-tunnel experiments of transition have been explained or at least clarified. In particular, several experimenters have observed "rope-like" waves in high-speed flows. Using the Flow Analysis Software Toolkit (FAST) developed at NAS, the Langley researchers were able to visualize the DNS results and to provide an explanation for the remarkably rope-like structures shown in the figure.



- The DNS data have been archived for future inquiries by other researchers. The fine detail of the flow field needed by transition modelers, for example, is available only from DNS.

Currently, the PSE-DNS approach is being extended to investigate transition on a highly swept supersonic wing, a configuration of immediate practical importance for the design of the HSCT.

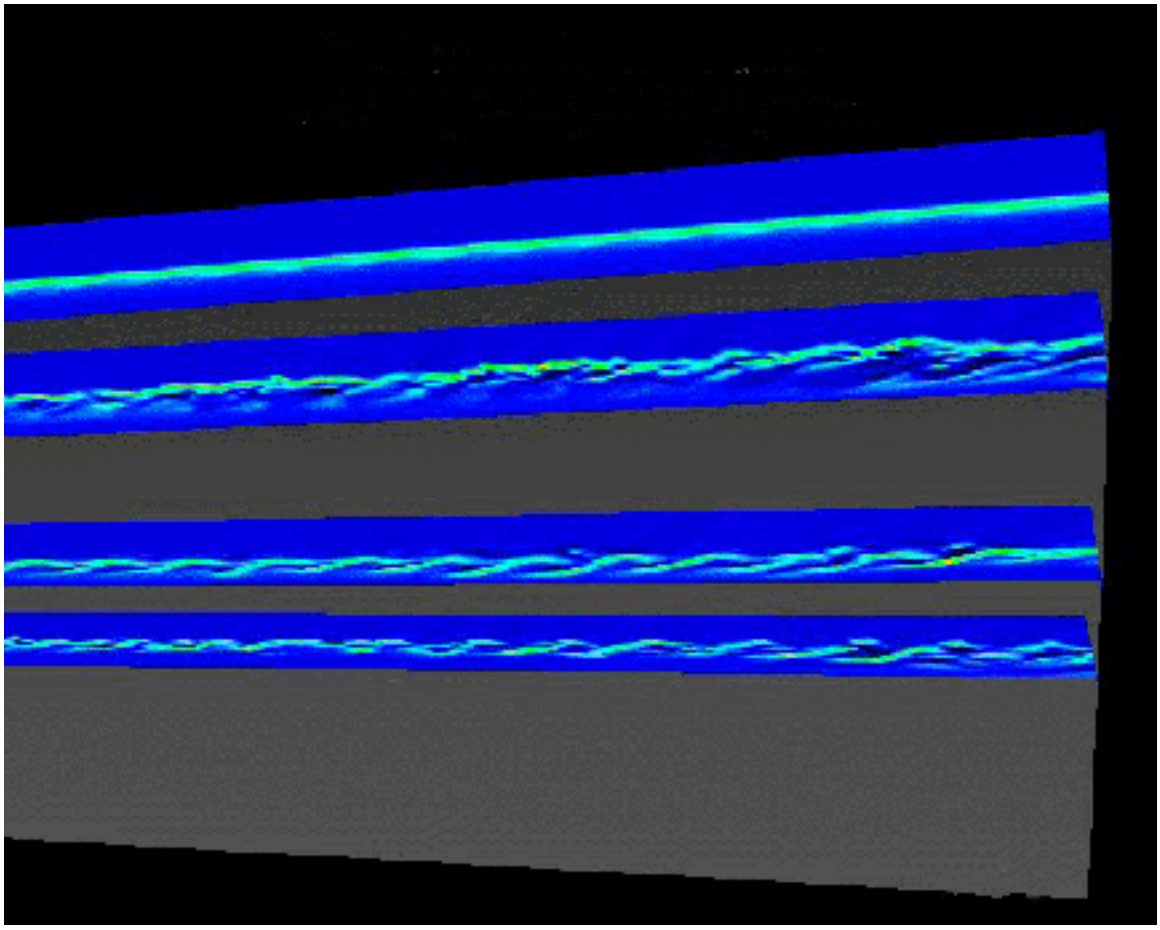
For more information, send email to Dave Pruett at **C.D.PRUETT@LARC.NASA.GOV**.

Dave Pruett holds a Ph.D. in Applied Mathematics from the University of Arizona. At the time this research was conducted, he was a senior scientist for Analytical Services and Materials Inc., Hampton, VA. Recently, he joined the Applied Sciences Program at the College of William and Mary, Williamsburg, VA, as a research scientist.

Chau-Lyan Chang holds a Ph.D. in Mechanical Engineering from Pennsylvania State University and is currently employed as a research scientist by High Technology Corp., Hampton, VA.

Both researchers gratefully acknowledge the former Theoretical Flow Physics Branch at NASA Langley, whose contractual support made this work possible.

Next Article | Contents | Main Menu | NAS Home



Visualization of direct numerical simulation results, which show a Mach 8 flow undergoing laminar-turbulent transition on the surface of a cone (a portion of which is shown in gray). The density gradient in selected planes perpendicular to the surface exhibits a rope-like appearance in regions where the flow remains essentially laminar. Other planes show nearly turbulent behavior.

---

to the Main Text

Metrics Forum Improves Quality of NAS Products and Services

by William Kramer

As part of NASA's Total Quality Management Program, in November 1993 the NAS Systems Division embarked on a new effort to measure and improve the quality and efficiency of its services and functions. Now in the second of three cycles, the NAS Metrics Forum, composed of representatives from almost every work group at NAS, meets weekly to share experiences learned from this effort.

The goals of the forum are to:

- exchange information about how each group measures its quality of service
- identify the most meaningful measurements and metrics
- set a baseline for each metric, defining what is "normal," from which deviations can be investigated
- periodically re-evaluate the performance of each area of service to determine appropriate levels for these metrics
- identify areas of excellence as well as areas for improvement

NAS has always used measurements and metrics to assess the quality of its services and to provide data for engineering and design decisions. In the past, however, each work group (generally associated with a subsystem or service function) determined the quantity and types of information to be tracked. Now, through the Metrics Forum, all NAS groups are working together to create standards for identifying and measuring quality improvement.

Three Cycles to Identify, Use Metrics

The first cycle of the forum concentrated on identifying what elements each group was already measuring and how measurements were made. The primary purpose of this cycle was to exchange information, experiences, and tools among groups. At this point, participants made no judgments about the meaningfulness of the measurements, but they did draw many conclusions and in-sights. They discovered that an immense amount of data was being recorded and that most of it was being reviewed in some way.

In the current (second) cycle, each work group proposes a set of metrics that are measurable in a meaningful way and that are "prime indicators" of its quality. Primary indicators are those measurements

that quickly and simply indicate the most about the health or quality of a system or function. Each group chooses only three to seven primary indicators, in order to focus on the most meaningful information.

Primary indicators are not represented by a single number, but may be a range of acceptable values within which to conclude that a system or service is working well.

The indicators may also be shown as rates of change instead of a fixed level. For example, indicators may measure the rate at which a new technology is evolving into production quality. Simply indicating the production level may not be a relevant measurement of a new system's functionality until much work is done; by using rates of improvement, progress and quality can be measured.

Setting these primary indicators does not mean that all other available data is ignored--it is still recorded and used for detailed problem solving, trend analysis, and engineering studies.

In the third cycle, groups will report on the actual collection of metrics and their experiences. It is expected that adjustments will be made to the identified metrics based on some period of observation. Adjustments may also be made to the ranges of "acceptable" levels for these metrics, based on experience and improvement.

Results Already Have an Impact

Although the forum activity is only partially completed, emerging results have already improved the quality of service at NAS. Some examples follow:

- The network support group developed a tool, called Xnet, which improved network monitoring, particularly on those systems that have not implemented Simple Network Management Protocol. Xnet allows the network support group to detect and correct marginal components in connections, hubs, and routers -- often *before* failures occur.

- The NAS user consultants regularly monitor the NAS and the Aeronautics Consolidated Supercomputer Facility CRAY C90 systems to identify codes that are candidates for tuning and optimization, and then offer code developers assistance with making these improvements. Metrics show significant improvements, in both system time and user CPU time, on individual codes that have been optimized by consultants. *(See figures)*

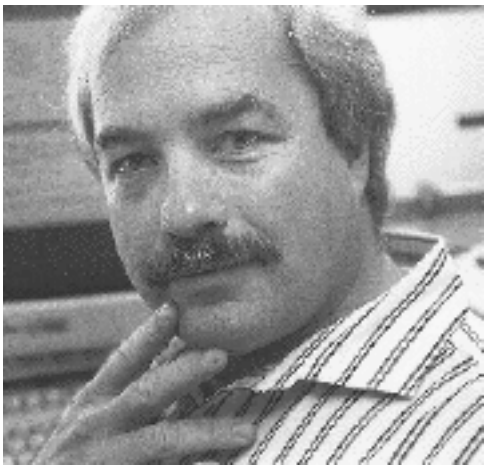
- As a precursor to the forum, a group of development and service staff members generated metrics specifications for the High Performance Computing and Communications Program testbed systems. These metrics are collected for all NAS parallel systems, as well as for the High Speed Processors, and provide a quantitative comparison for the testbed systems' progress toward becoming production supercomputers.

Some of these metrics are already being implemented in automated monitoring tools under the NAS operational tools project, in order to provide proactive problem detection.

Forum representatives have found that choosing primary indicators and setting baselines is difficult work. They also quickly observed that it's easier to gather, quantify, and understand system metrics than service metrics. More emphasis is now being placed on ways to measure the quality that a system delivers to end users, rather than what it seems to be doing from the technician's viewpoint. Forum participants have also observed that many groups are recording and presenting similar data in different ways. Work has begun to provide tools for these common areas--such as problem-tracking data and evaluations of NAS training classes.

For More Information

Check future issues of *NAS News* for progress on the Metrics Forum. For more information, look at the Metrics page or send email to **kramer@nas.nasa.gov**.



William Kramer created and has been the driving force behind the NAS Metrics Forum. He has been a branch chief at NAS for the past six years.

Next Article | Contents | Main Menu | NAS Home

Figure 1. Chart shows the improvement in the megaflop per second rate on four user codes after optimization by the NAS user services group. The group chose this rate as a primary indicator because it reflects the ability of the code to take advantage of special features in the Cray architecture.

Figure 2. Graph displays the decreases in CPU time on the same four user codes (shown in Figure 1) after optimization by NAS user services consultants. The third section of the chart shows the overall decrease in CPU time, which the consultants chose as a primary indicator because it quantifies the overall "savings" made by optimizing code, including scalar, vector, and I/O optimization. As shown in the chart, the consultants' goal--to decrease overall CPU by 20 percent or more--has been met or exceeded for these four codes.

Back to the Main Text

New Program at NAS Aids Large File Transfers

by Elisabeth Wechsler

NAS has developed a completely new, specialized software program that aids in the transfer of large solution files between supercomputers and workstations or mass storage systems.

The Asynchronous Remote Copy Program (ARCP) is based on some features of File Transfer Program (FTP) and Remote Copy Program (RCP) but is not intended to replace them, according to Jude George, of the NAS wide-area network development group.

More `Reliable' for Large Files

"Scientists want a better mechanism for moving their solution files back and forth between the Crays, workstations, and mass storage," George said, adding that although FTP and RCP usually get the job done, these programs are not totally reliable for 50-megabyte files. "FTP and RCP are good for lots of other things, but ARCP is written specifically for the supercomputing community."

The typical scenario addressed by ARCP is for a scientist to create a solution file on a Cray, then transfer it to a workstation to visualize the results or send it directly to mass storage, he explained.

Since transfers of large files are usually very time consuming (sometimes requiring several hours), they can be aborted more easily -- either because the host reboots, or the network connection "drops out" or becomes so error-prone that the connection is closed and the file transfer is stopped, George said. While emphasizing that this does not happen often, he said, "it's still a big deal to that researcher" when it does occur.

Saves Supercomputing Resources

If the transfer is stopped in such a situation, the scientist risks losing the file altogether. "Because supercomputing resources are at such a premium (with limited disk space shared by lots of people), there is no economical way to avoid purging files that don't complete the transfer," George explained. Thus, the scientist would need to run the solution file again, thereby compounding the resource problem.

ARCP starts the transfer and monitors it -- intervening when necessary -- giving scientists more reliability during the process and expanding the available time on the Crays. "ARCP either gets the file transferred or you're notified," he said, adding that the program "tries very hard to get the file across."

Reducing Instruction Buffer Fetches

by George Myers

The Cray hardware performance monitor (hpm) provides a wealth of information about job performance. In addition to quantifying megaflops (millions of floating point operations per second), hpm identifies problem areas in code. For instance, the value of the field containing instruction buffer fetches per second indicates the efficiency of program execution; a value greater than .15 million may indicate some inefficiency.

Why? Each CPU has a stack, or buffer, which stores a block of instructions. Each time the flow of execution takes the CPU out of this stack, a new instruction block must be read from memory. Reading memory is expensive compared to reading an instruction from the stack. This expense is part of a job's system CPU time; so, reducing the frequency of instruction buffer fetches reduces system CPU time and overall CPU time. Frequent instruction buffer fetches also increase the possibility of memory contention (bank conflicts), further reducing program efficiency.

Several common reasons for a high number of instruction buffer fetches per second are: "Spaghetti" code--code that jumps around a lot--which requires an algorithm change to correct; making subroutine calls in major loops in the program; an instruction buffer boundary occurs inside a loop, or the length of the loop length requires more than one block of instructions. The latter two causes are addressed here. **`Inlining' Solves Call Problems** A simple solution, "inlining," solves the subroutine call problem, and the payback is greater than simply reducing instruction buffer fetches. Each subroutine call requires an instruction buffer fetch and overhead is added to set up the parameter list for the subroutine being called. Inlining corrects both of these "costs" by placing the source from the subroutine within the calling routine before compiling that code segment. Besides saving CPU time, inlining frequently allows loops that couldn't vectorize (due to the presence of a subroutine call) to do so. Take care to weigh the performance benefits of lining against the cost--an increase in the size of the executable file, which is dependant on the size of the routines being inlined and the number of different places the routine is called.

Users can direct the compiler to attempt to inline subroutines; two of several methods follow.

Inlining can be done by the compiler (cft77) or the preprocessor (fpp). Cft77 follows some basic built-in rules when attempting to inline code, while fpp allows the user to determine some of these rules.

At times, it may be convenient to put all subroutines to be inlined in a single file and then use the compiler directives to specify that file. The cf77 command-line for this is:

```
cf77 -Wf"-I infile" program.f
```

where **-Wf"-I infile**" directs the compiler to inline all subroutines contained in file *infile*.

Another method is to delineate certain constraints such as the number of levels of subroutines to inline and the maximum subroutine size, shown in the line:

```
cf77 -Zv -Wd"-J2 -M1000 -S ." program.f
```

where **-Zv** activates the preprocessor, fpp, and where -Wd"-J2 -M1000 -S ." directs fpp to try to inline two levels of subroutines (-J2) found in the current directory (-S .) of 1000 lines or less (-M1000) where they are called from program.f.

Another Solution: `Loop Alignment'

Cray provides a solution to correct an instruction block boundary occurring within loops. With "loop alignment," the compiler tries to align a block of instructions with the beginning of a loop so that the entire loop is contained in the instruction stack. Activate this feature with the command line:

```
cf77 -Wf"-o loopalign" program.f
```

where **-Wf"-o loopalign"** instructs cft77 to align loops on instruction buffer boundaries.

These two techniques can produce significant improvement in program execution, reducing the CPU time consumed by programs, and saving time and money.

For more information, see the man pages for cf77, cft77, and fpp, or use the docview command to get the Cray manual *CF77 Commands and Directives*. Contact NAS User Services at (415) 604-4444 or send email to **nashelp@nas.nasa.gov**.

Next Article | Contents | Main Menu | NAS Home

Chancellor's Background and Experience Support Diversity at NAS

by Jill Dunbar

As Deputy Chief of the NAS Systems Division, and currently serving as Acting Chief, Marisa Chancellor relies on a wide range of skills--including a working knowledge of computer hardware and software, technical project management, and government-contractor relations--to have an effect on the success of the organization.

"A systems and industry background contribute to my knowledge of what [the aerospace] industry's expectations are," Chancellor said. And she knows both sides of government and contractor issues: Before joining government service, Chancellor worked for Control Data Corp. (CDC), which served as a contractor at NASA Ames Research Center.

Her Job: To Ensure NAS Quality

Chancellor likens her job as Deputy Chief to that of a corporate chief operations officer, handling a variety of day-to-day functions. Her primary goal of "making sure that NAS maintains a high level of quality and is organized better around users' changing needs" requires a flexible attitude.

She explained that, on the other hand, the job of Division Chief is like a chief executive officer who must have vision, deal with high-level industry representatives, communicate effectively with NASA headquarters, and so on. "That's what Dave [Cooper] deals with, and he does it well," she said. Chancellor noted that Cooper began a mentoring program with her and will continue in that role, and added, "I find that help invaluable."

Chancellor has been Acting Chief of NAS since Cooper's appointment as head of the newly created Ames Information Systems Directorate (designated as Code I; see the related front-page article in the September-October issue of *NAS News*). Her focus in this position is to ensure that NAS continues to run smoothly while the search for a permanent chief is conducted. Once that position is filled--a task expected to take up to a year--Chancellor will resume her duties as deputy, which are currently being handled by Eric Hibbard, who leads a computer graphics-scientific visualization group in the NAS Data Analysis Branch.

Sees Changes as `Very Exciting'

Chancellor views the recent organizational changes at Ames as "very exciting" and said that "everyone [on the Code I management team]-- especially the core transition team--is excited about where this organization can go." She added that Cooper, Code I Deputy Director Henry Lum, and others are making a real effort to create a cohesive, strategic plan.

According to Chancellor, the Code I management team is working specifically on ways to motivate, recognize, and reward individuals as well as teams for outstanding contributions.

She noted that in the new organization more value will be placed on staff contributions in the service and support areas. "Some people have an attitude that support is less valuable than producing scientific and technical papers." But now, she continued, "Code I has an opportunity to promote new ways of thinking."

Code I will also focus on minimizing time-consuming administrative tasks, emphasizing, for example, a "paperless" office environment. Chancellor pointed out that goals of cost-effectiveness and efficiency allow spending more discretionary funds for new technology in specific fields, as well as devoting more time to innovative projects, as opposed to concentrating on "firefighting."

Chancellor recognizes that change is hard to adjust to. She is aware of the stress that reorganization causes staff members and stated that "good work is recognized and valued anywhere in an organization."

Brings an Employee Perspective

Chancellor describes herself as a "people person." She believes she understands what motivates people, has a knack for establishing good working relationships, and empathizes with those who deal with the ups and downs associated with all projects.

Her strategies include focusing on common goals; fostering closer, tight-knit work groups; and promoting individual achievement. She believes in assigning projects that will enhance individuals' capabilities and professional growth.

From her five years working as a government contractor, she feels that she brings an employee perspective to the deputy position, and understands how someone might relate to being on the receiving end of management decisions.

Experience Focused on Computing

Chancellor was appointed to her current position at NAS in April 1993 (having been Acting Deputy since November '92), when former Deputy Chief Ron Deiss took responsibility for designing the Aeronautics Consolidated Supercomputing Facility project.

Before her promotion to deputy, Chancellor was the NAS workstation development group leader. She

was responsible for overseeing the development and implementation of scientific visualization software, working with scientists to determine appropriate software features and scheduling program phases.

While at CDC, she was the on-site analyst for the CDC Cyber 205 supercomputer and the analyst-in-charge of the ETA-10 supercomputer. In the latter role she supervised a team of four analysts and three engineers, handled all test phases and day-to-day system operations, and worked with UNIX software internals and network protocols.

She also applied her knowledge of Fortran to assist users in compiling and running large computational fluid dynamics programs.

Chancellor earned her Bachelor's degree in computer science and mathematics at the University of California, Davis.

Represents an Alternative Viewpoint

"I'm proud of the fact that I can represent views not normally held by [Ames Research] Center managers. They're not radical views--just different," she said.

For example, her opinion differs from the so-called politically correct viewpoint on diversification in management. As a native Californian who was raised in a middle-class community, Chancellor doesn't perceive that she has "ever been on the receiving end of discrimination." But, as a woman who qualifies for (but doesn't claim) a protected-class status, she is "very much aware that there may be other factors besides job skills" taken into account when considering job candidates. She has expressed to Ames executives her view that, as a manager, she wants to hire the best qualified person--regardless of their status--and that, as a candidate, she would want to *be* hired based on her ability to do the job.

Judged `Top-notch' Student Essays

Other recent contributions she is proud to have been part of include being on a panel to evaluate an annual technical paper contest sponsored by the Ames Advisory Committee for Women, as well as judging technical essays for a contest sponsored by the Ames Educational Programs Office, in which high-school seniors competed for a $2,500 college scholarship.

"It was nice to get a perspective on young people that's different from the typical way society sees teenagers," she said, noting that the student essays were very impressive and adding, "It's heartening see a large group of kids out there who are really top-notch people."

Marisa Chancellor, NAS Deputy Chief, also serves as Executive Editor of NAS News and encourages readers to voice their opinions about the publication.

Next Article | Contents | Main Menu | NAS Home

NAS News, Nov - Dec 1994
Vol. 2, No. 6

Executive Editor: Marisa Chancellor

Editor: Jill Dunbar

Senior Writer: Elisabeth Wechsler

Contributing Writers: Chau-Lyan Chang, Samuel Fineberg, William Kramer, George Myers, C. David Pruett

Image Enhancements: Chris Gong

Other Contributors: David Bailey, Chris Beaumont, Jeff Becker, Jim Craw, Leonardo Dagum, Pat Elson, Jude George, Steve Harding, Toby Harness, Ed Hook, Bob Hirsch, Richie Jacobovits, Tom Lasinski, Jim McCabe, John Musch, Terry Nelson, Chuck Niggley, Bill Nitzberg, Bill Saphir, Leigh Ann Tanner, Arsi Vaziri, Howard Walter, Tom Woodrow

Editorial Board: Marisa Chancellor, Jill Dunbar, Chris Gong, Creon Levit, Serge Polevitzky, Pamela Walatka, Elisabeth Wechsler, Rita Williams

NAS NEWS
Volume 2, Number 6 November – December 1994

MPIRUN: A Loader for Multidisciplinary and Multizonal MPI Applications

by Samuel Fineberg

MPIRUN, a NAS-developed utility, allows users to create multidisciplinary and multizonal application programs that are portable to any parallel computer or to clusters of workstations. MPIRUN works with both C and Fortran and is currently available on all NAS parallel systems.

Programs Important in Aerosciences

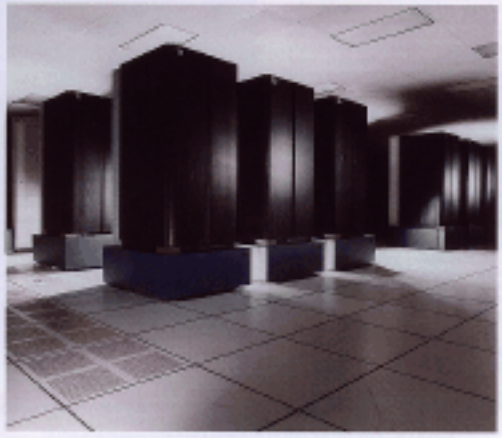
Multidisciplinary and multizonal applications are an important class of programs in the area of computational aerosciences. In these codes, two or more distinct parallel applications or copies of a single application are used to model a single problem. These parallel applications are usually implemented using message passing. All message-passing libraries support simple applications consisting of a single program running on many processors (often called SPMD). However, support is lacking for applications where different programs are running on each processor. Unfortunately, this is the most straightforward way to implement multidisciplinary and multizonal applications, utilizing a separate program for each discipline or zone.

In the past, these applications were implemented using extensions to vendor-supplied message-passing libraries, for example, the intercube library for the Intel iPSC/860 and the Iblap library for the Intel Paragon. However, programs written with these libraries are not portable.

There are many portable message-passing libraries, some of which support the communications requirements of multidisciplinary and multizonal applications. One well-known portable message-passing library is Parallel Virtual Machine (PVM) from the University of

Continued on page 6

THIS ISSUE



The newest parallel supercomputer system at NAS, a 159-node IBM SP2 configuration called "Babbage," consists of 16 servers and 8 vendor nodes. The system was installed in September and serves approximately 300 user accounts. The three-year Cooperative Research Agreement is a partnership with industry, government, and university representatives.

IBM SP2 Delivers Performance, Challenges to Consortium Partners

by Elisabeth Wechsler

Initial configurations of the IBM SP2 system passed acceptance tests and the NASA Computational Aerosciences (CAS)-funded Cooperative Research Agreement (CRA) remains on schedule. However, significant challenges await the consortium in the months ahead for both hardware and software development, according to various members of the project team at NAS and IBM.

NAS personnel have expressed considerable satisfaction with the performance that has been achieved to date (*see sidebar article on page 2*). Noted below, some problems remain. The principal area of concern is system software, including the need for dynamic job management and parallel file systems support. IBM and NAS Systems Division personnel are looking into these areas on behalf of the consortium.

For background information about the CRA, see the article entitled, "IBM Consortium Awarded $22.4 Million for CAS Research Project," in the July–August issue of NAS News.

"We expected some shortcomings in system software and hope to work with IBM to make the system more usable to researchers," said David Bailey, who coordinates the CRA research project at NAS.

Richie Jacobowitz, IBM program manager for the CRA, said: "We've seen the benefits of collaboration and IBM is really happy with the way things are going." He continued, "The implementation of the acceptance test was much more rigorous than we expected and we're glad we've been able to live up to the expectations NAS has set." He explained that NAS "created a script that ran the evaluation test suite around the clock. The acceptance test was very much a stress test because the test script ran the SP2 at peak performance rates for nearly two weeks."

Making a 'True' Collaboration

"We're trying to make this a true collaboration," said Toby Harness, of the NAS parallel systems group. "Our goal is to make IBM do well—there's no 'vendor' involved because it's a CRA."

Chuck Niggley, of the NAS parallel systems science support group, echoed that point of view. "We've established a support within various IBM development groups so we can have contributions/ access to some of their future design specifications. They are beginning to welcome our input."

Each week, 15–35 NAS staff and researchers attend a teleconference, joined by 2–3 IBM

Continued on page 2